**软件需求分析**

题目：基于机器学习算法的影视数据分析系统

导师：袁宇丹

学生：邓高登 崔力辉 王世卿

目录

[1. 引言 3](#_Toc499050587)

[1.1 编写目的 3](#_Toc499050588)

[1.2 项目背景 3](#_Toc499050589)

[2. 任务概述 3](#_Toc499050590)

[2.1 开发目标 3](#_Toc499050591)

[2.2 开发环境 4](#_Toc499050592)

[2.3 技术要求 4](#_Toc499050593)

[2.4 用户界面要求 4](#_Toc499050594)

[3. 项目具体方案 5](#_Toc499050595)

[3.1 影视数据部分 5](#_Toc499050596)

[3.2 机器学习部分 5](#_Toc499050597)

[4. 系统架构 5](#_Toc499050598)

[5. 用例图 6](#_Toc499050599)

[6. 数据字典 7](#_Toc499050600)

[7. 参考文献 8](#_Toc499050601)

基于机器学习算法的影视数据分析系统

# 引言

## 编写目的

本文从项目背景、项目目标、项目方案、项目实现、项目测试等方面对影视数据采集与分析系统进行了详细的叙述。在设计与实现该系统的实际工作中，我们将从数据采集、数据清洗、数据分析和数据可视化四个方面深度挖掘影视数据所蕴含的价值，为投资者投资影视市场提供合理建议。

## 项目背景

随着我国网民规模的增大，影视市场的快速壮大，视频网站的快速发展以及新型社交媒体的助力，我国的影视市场产生并积累了巨大的有效和有用的数据，在信息时代，这无疑是一笔巨大的财富。随着机器学习、自然语言处理等技术的发展，对于数据的价值挖掘和利用越来越受到人们的重视。

从近些年持续火爆的影视票房来看，我国影视市场无疑具有很大的发展前景。如何在影视市场中掌握先机，立于领先地位，是很多人特别是电影投资商关注的问题。针对这样的背景，本文提出了基于机器学习算法的影视数据分析系统，以期在对影视市场积累的数据进行充分利用的基础上，能够给予投资者投资影视市场的一些合理建议。

# 任务概述

## 开发目标

对软件进行功能需求分析、性能分析、界面需求分析、可靠性需求分析以及时效性需求分析等其他方面的需求分析。

## 开发环境

(1) 运行环境：linux

(2) 数据存储：mongoDB

(3) 编程语言：Python

(4) 算法库：Jieba, Nltk, Numpy, scikit-learn, snownlp

## 技术要求

软件需求分析阶段的技术要求如下：

（1）软件需求规格要说明对应软件的主要功能、性能、技术指标进行定义，其内容应全面、可检查。

（2）项目开发计划中应给出阶段评审以及配置管理计划。

（3）软件需求规格中说明书要正确而恰当的定义软件的功能、性能等所有软件需求。

（4）要求编制的软件需求规格要具有：完整性、明确性、一致性、可验证性以及可测试性、易修改、包含软件需求关键功能、需求基于运行环境、描述软件应发生的事件和不应该发生的事件。

## 用户界面要求

其用户界面要求为包括简洁、易懂、易用、美观，具备一定的兼容性；同时符合用户的操作习惯。

# 项目具体方案

## 影视数据部分

1. Requests、Beautifulsoup：实现影视数据爬取
2. mongoDB：实现影视数据存储

## 机器学习部分

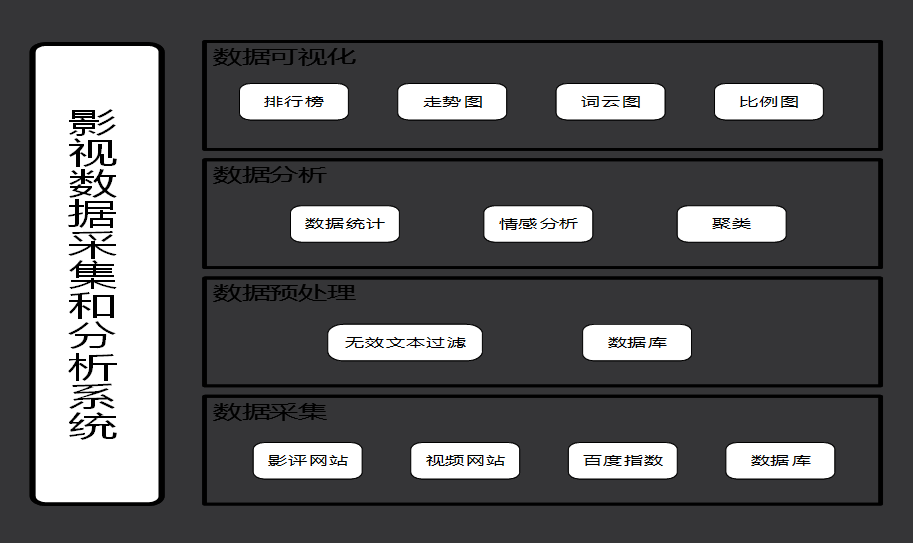
（1）Jieba

（2）scikit-learn

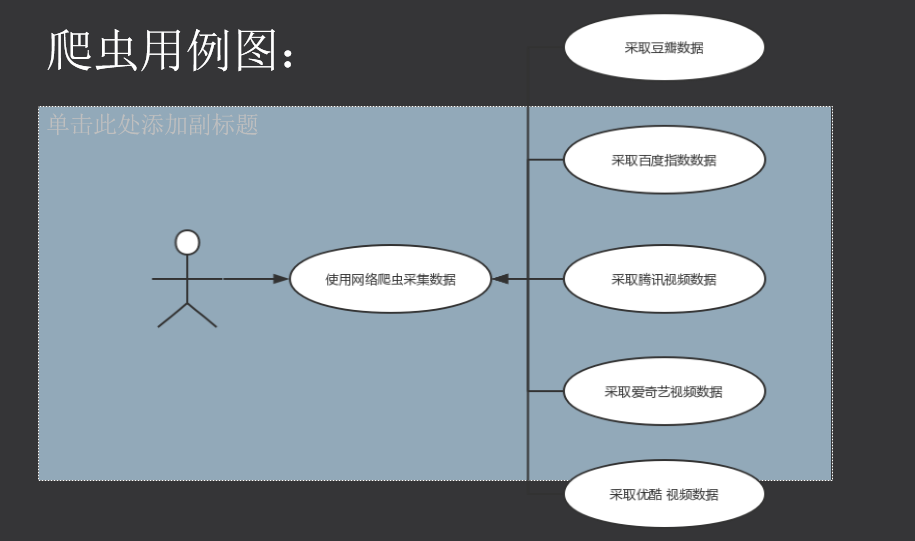
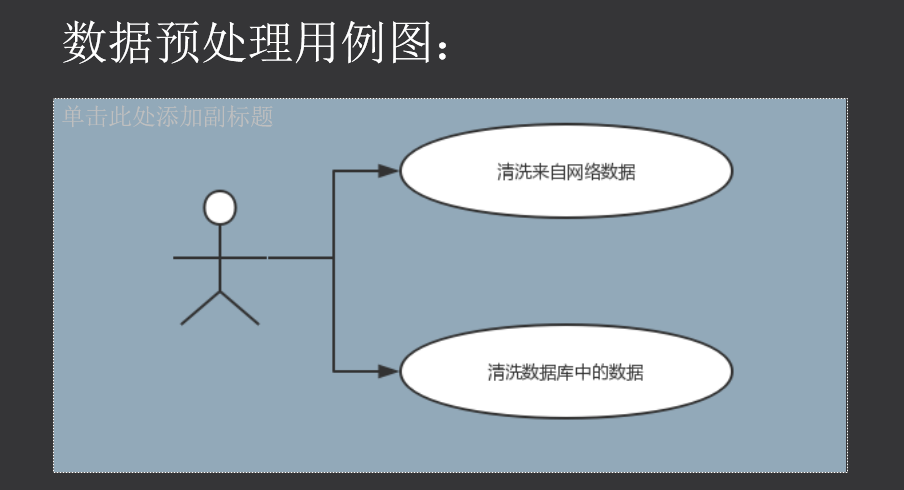
（3）snownlp

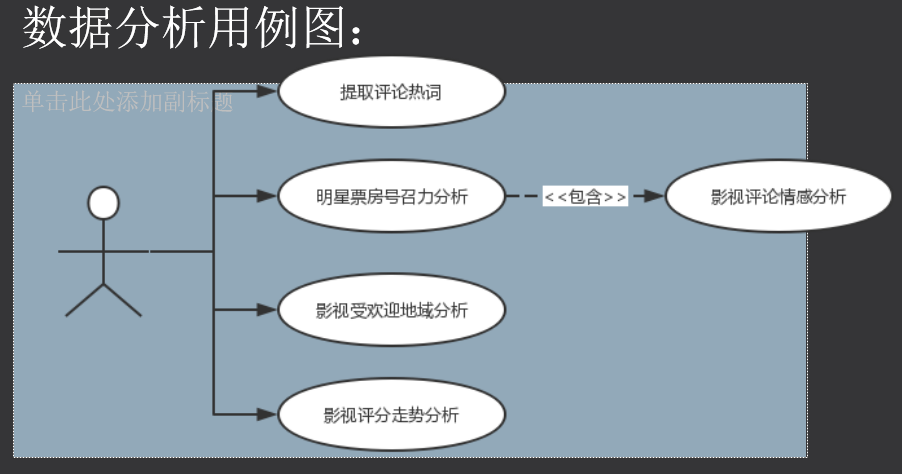
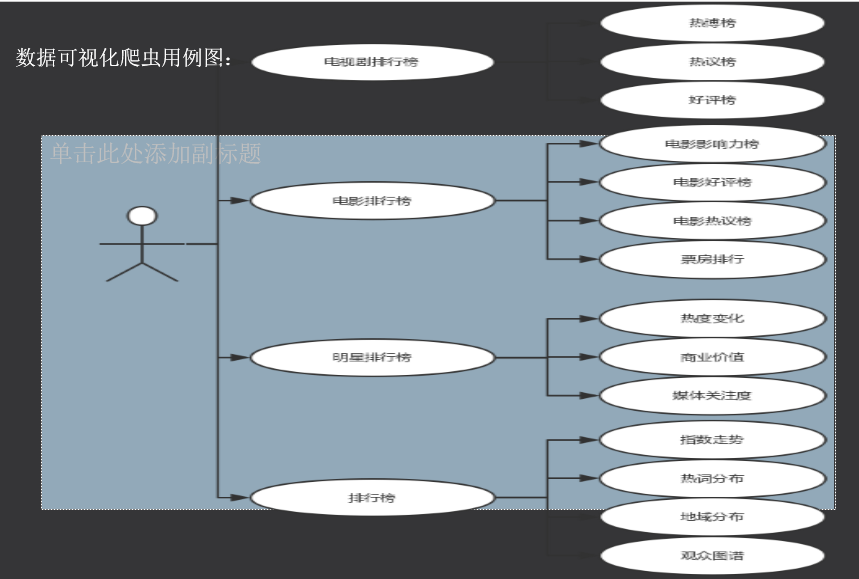
（4）Nltk

# 系统架构



# 用例图

# 数据字典

（1）影视播放量表

描述：影视播放量集合

定义：影视播放量表=节目名+本期名+本期日期+本期播放量+本期链接

（2）影视评分表

描述：影视评分集合

定义：影视播放量表=节目名+本期名+本期日期+本期评分+本期链接

（3）电影推荐表

描述：电影推荐表集合

定义：电影推荐表=用户ID+用户购买电影记录+推荐电影类型

# 参考文献

[1]王志梅，杨帆.基于Hebbian一致性学习的推荐算法[J].计算机研究与应用，2006,42(36):100-113.

[2]吴彦, 沈洁，顾天竺等.协同过滤推荐系统中数据稀疏问题的解决